

# The Gaussian Channel with Noisy Feedback: Improving Reliability via Interaction

Assaf Ben-Yishai and Ofer Shayevitz

**Abstract**—Consider a pair of terminals connected by two independent (feedforward and feedback) *Additive White Gaussian Noise* (AWGN) channels, and limited by individual power constraints. The first terminal would like to reliably send information to the second terminal at a given rate. While the reliability in the cases of no feedback and of noiseless feedback is well studied, not much is known about the case of noisy feedback. In this work, we present an interactive scheme that significantly improves the reliability relative to the no-feedback setting, whenever the feedback Signal to Noise Ratio (SNR) is sufficiently larger than the feedforward SNR. The scheme combines Schalkwijk-Kailath (S-K) coding and modulo-lattice analog transmission.

## I. INTRODUCTION

Feedback cannot improve the capacity of point-to-point memoryless channels [1]. Nevertheless, noiseless feedback can significantly simplify the transmission schemes and improve the error probability performance, see e.g. [2]–[4]. These elegant schemes fail however in the presence of arbitrarily small feedback noise, rendering them grossly impractical. This fact has been initially observed in [2] for the AWGN channel, and further strengthened in [5].

In a previous work [6] we presented a variation of the noiseless-feedback AWGN S-K scheme [2], extending it to the case of noisy feedback. The scheme was based on the following observation: In each round, the receiver has some estimate of the message, and the transmitter needs to learn the associated estimation error in order to proceed. This estimation error can be conveyed in a power-efficient manner by using the knowledge of the message at the transmitter as side-information. The main focus of [6] was on the simplicity of the scheme in a fixed error probability regime, and side information was used by applying scalar modulo operations. This resulted in a major improvement of the capacity-gap in a relatively small number of rounds.

The focus of this work is on the virtues of noisy feedback for increasing reliability. To that end, an asymptotic generalization of the scheme in [6] is introduced, applying the S-K scheme over blocks and replacing the scalar modulo with multi-dimensional lattice modulo, as well as replacing Pulse Amplitude Modulation (PAM) used in [6] with a block code. An asymptotic error analysis is provided, using the Poltyrev exponent to account for modulo aliasing errors, and channel coding error exponents to account for the error of the block code. The resulting error exponent is computed and shown to

surpass the sphere-packing bound of the feedforward channel for a wide range of rates and SNR settings.

In [5], [7], the authors analyzed the reliability function of the AWGN at zero rate for noisy *passive* feedback, i.e. where the channel outputs are fed back without any processing. In [8], which is closer to our interactive setting, the reliability function of the AWGN at zero rate (two messages) with noisy *active* feedback has been considered. Specifically, it was shown that active feedback roughly quadruples the error exponent relative to passive feedback. The achievability result of [8] is better than ours at zero rate.

## II. PRELIMINARIES

We write  $\log$  for base 2 logarithm, and  $\ln$  for the natural logarithm. We use the vector notation  $x^n \stackrel{\text{def}}{=} [x_1, \dots, x_n]$  and boldface letters such as  $\mathbf{x}$  to indicate vectors of size  $N$ . We write  $a_n \geq b_n$  to mean  $\liminf_{n \rightarrow \infty} \frac{1}{n} \ln \left( \frac{a_n}{b_n} \right) \geq 0$ , and similarly define  $\leq$  and  $\doteq$ .

### A. Lattice Properties

- (i) We denote a lattice of dimension  $N$  by  $\Lambda = G \cdot \mathbb{Z}^N$  where  $G$  is the generating matrix.
- (ii)  $V(\Lambda) = |\det(G)|$  is the lattice cell volume.
- (iii) We denote the nearest neighbor quantization of  $\mathbf{x}$  to the lattice  $\Lambda$  by  $\mathbb{Q}_\Lambda[\mathbf{x}]$ .
- (iv) We denote the fundamental Voronoi cell  $\mathcal{V}_0 = \{\mathbf{x} : \mathbb{Q}_\Lambda[\mathbf{x}] = \mathbf{0}\}$ .
- (v) Modulo  $\Lambda$  is defined as  $\mathbb{M}_\Lambda[\mathbf{x}] \stackrel{\text{def}}{=} \mathbf{x} - \mathbb{Q}_\Lambda[\mathbf{x}]$ .
- (vi)  $\mathbb{M}_\Lambda[\cdot]$  satisfies the *distributive law*:  $\mathbb{M}_\Lambda[\mathbb{M}_\Lambda[\mathbf{x}] + \mathbf{y}] = \mathbb{M}_\Lambda[\mathbf{x} + \mathbf{y}]$ .
- (vii) The volume to noise ratio (VNR) of a lattice in the presence of AWGN with variance  $\sigma^2$  is  $\mu \stackrel{\text{def}}{=} V^{2/N}(\Lambda)/\sigma^2$ .
- (viii) The normalized second moment of a lattice  $\Lambda$  is  $G(\Lambda) \stackrel{\text{def}}{=} \sigma^2(\Lambda)/V^{2/N}(\Lambda)$ , where  $\sigma^2(\Lambda) = \frac{1}{N} \mathbb{E}(\|\mathbf{U}\|^2)$  and  $\mathbf{U}$  is uniformly distributed on  $\mathcal{V}_0$ .

### B. Joint Source Channel Coding (JSCC)

It is well known [9] that when a Gaussian source is conveyed over AWGN channel and quadratic distortion measure, analog transmission obtains the optimal distortion with minimal delay. The transmitter merely has to scale the source  $Q$  to the channel input power constraint, and the receiver merely has to multiply by the appropriate Wiener coefficient in order to obtain the optimal linear estimate. This solution is a simple case of joint source channel coding (JSCC).

If side information related to the source  $Q$  is present at the receiver the problem is known as the Wyner-Ziv problem [10].

The authors are with the Department of EE-Systems, Tel Aviv University, Tel Aviv, Israel {assafbster@gmail.com, ofersha@eng.tau.ac.il}. The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. 639573, and from the Israel Science Foundation under grant agreement no. 1367/14.

Kochman and Zamir [11] gave the solution of JSCC with side information (related to the source and the channel) over an AWGN with quadratic distortion measure. They used analog transmission as [9] in conjunction with dithered modulo lattice operations that take care of the side information.

Let us quickly quote survey their result in the case of side information only at the source. The source vector to be conveyed is  $\mathbf{Q} + \mathbf{J}$  where the destination has  $\mathbf{J}$  as side information. The channel is AWGN with input  $\mathbf{X}$  noise  $\mathbf{Z}$  and output  $\mathbf{Y}$ , i.e.  $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$ . The transmitter sends:

$$\mathbf{X} = \mathbb{M}_\Lambda [\beta(\mathbf{J} + \mathbf{Q}) + \mathbf{V}]$$

Where  $\mathbf{V}$  is the dither vector uniformly distributed on  $\mathcal{V}_0$  the basic Voronoi cell of  $\Lambda$  and commonly known at the transmitter and receiver. The receiver first calculates the temporary variable  $\mathbf{T}$  as follows:

$$\mathbf{T} \stackrel{\text{def}}{=} \alpha_C \mathbf{Y} - \mathbf{V} - \beta \mathbf{J} = \mathbf{X} + \mathbf{Z}_{\text{eq}} - \mathbf{V} - \beta \mathbf{J}$$

where the second transition is pedestrian by the definition of the equivalent noise  $\mathbf{Z}_{\text{eq}}$ :

$$\mathbf{Z}_{\text{eq}} \stackrel{\text{def}}{=} -(1 - \alpha_C) \mathbf{X} + \alpha_C \mathbf{Z}$$

The receiver now applies another modulo operation on  $\mathbf{T}$  obtaining  $\mathbf{U}$  as follows:

$$\mathbf{U} \stackrel{\text{def}}{=} \mathbb{M}_\Lambda [\mathbf{T}] = \mathbb{M}_\Lambda [\beta \mathbf{Q} + \mathbf{Z}_{\text{eq}}]$$

where the second transition is due to the distributive law on  $\mathbb{M}_\Lambda [\cdot]$ . Now, if  $\beta \mathbf{Q} + \mathbf{Z}_{\text{eq}} \in \mathcal{V}_0$ , then  $\mathbf{U} = \beta \mathbf{Q} + \mathbf{Z}_{\text{eq}}$ . We show in the sequel that by appropriate parameter settings, the probability of the complementary event can be made exponentially small with respect to the lattice dimension  $N$ .

In [11] a linear estimate of  $\mathbf{Q}$ :  $\hat{\mathbf{Q}} = \frac{\alpha_S}{\beta} \mathbf{U}$  was obtained. However, in our scheme  $\frac{\alpha_S}{\beta}$  naturally cancels out rendering the setting  $\alpha_S$  immaterial. We note that setting  $\alpha_C < 1$  is common practice in many lattice problems, improving performance in lower SNRs and making  $\mathbf{Z}_{\text{eq}}$  non-Gaussian (which usually improves the error probability). For clarity of exposition we use in this work only  $\alpha_C = 1$ .

### C. The Schalkwijk-Kailath (S-K) Scheme

The famous S-K scheme[2] for capacity achieving communication over AWGN can be interpreted using JSCC tools. The classic scheme encodes the message  $W$  into a message point  $\Theta$  using single dimensional modulation. At the end of the first step,  $k = 1$ , Terminal B sets its estimate of  $\Theta$  to be  $\hat{\Theta}_1 = Y_1$ . In consequent steps  $k$ , Terminal B feeds back  $\hat{\Theta}_k$  to Terminal A. At step  $k + 1$ , Terminal A extracts the estimation error  $\varepsilon_k = \hat{\Theta}_k - \Theta$  and conveys it to Terminal B by JSCC. Namely, Terminal A sends  $X_k = \alpha_{k+1} \varepsilon_k$  where  $\alpha_{k+1}$  is set so that to meet the channel input power constraint and Terminal B linearly estimates  $\hat{\varepsilon}_k = \beta_{k+1} Y_{k+1}$  where  $\beta_{k+1}$  is set so that to minimize the *Mean Squared Error* (MSE). Having  $\hat{\varepsilon}_k$ , Terminal B now advances its estimate by  $\hat{\Theta}_{k+1} = \hat{\Theta}_k + \hat{\varepsilon}_k$ . Finally, at step  $K$ , Terminal B decodes  $W$  from  $\hat{\Theta}_K$ .

For the sake of analysis it is convenient to observe the series of channels from  $\Theta$  to  $\hat{\Theta}_k$ . These are Gaussian channels whose noise variance is  $\sigma^2$  at  $k = 1$ , and it is easy to see that optimizing over  $\alpha_k$  and  $\beta_k$  reduces the noise variance by

$1 + \text{SNR}$  at every step. So, at the final step  $K$  we have a channel whose SNR is  $\text{SNR}(1 + \text{SNR})^{K-1}$ . At this step it can be shown that mapping  $W$  into  $\Theta$  using PAM and giving a Gaussian analysis of the error probability, can yield a rate arbitrarily close to the channel capacity by taking a sufficiently large  $K$ .

### D. Error Exponents

Consider the case where a lattice point  $\mathbf{X} \in \Lambda$  is sent over an AWGN channel  $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$  and the decoder estimates  $\hat{\mathbf{X}}(\mathbf{Y})$  according to an *Maximum Likelihood* (ML) decoding rule. Then there exist lattices whose probability of decoding error is exponentially upper bounded by  $\Pr(\hat{\mathbf{X}}(\mathbf{Y}) \neq \mathbf{X}) \leq e^{-N E_p(\frac{\mu}{2\pi\epsilon})}$ , where  $\mu$  is the VNR w.r.t the lattice and the channel noise variance and  $E_p(\cdot)$  is the Poltrev error exponent given by [12], [13]:

$$E_p(x) = \begin{cases} \frac{1}{2}(x - 1 - \ln(x)) & \text{if } 1 < x \leq 2 \\ \frac{1}{2}(\ln(x) + \ln(\frac{\epsilon}{4})) & \text{if } 2 < x \leq 4 \\ \frac{1}{8}x & \text{if } x > 4 \end{cases}$$

For channel coding over AWGN with SNR and with rate  $R$ , there exist block codes of length  $N$  whose average error probability (averaged over the messages) under ML decoding is exponentially upper bounded by  $\Pr(\hat{\mathbf{X}}(\mathbf{Y}) \neq \mathbf{X}) \leq e^{-N E_r(R)}$  where  $E_r(\text{SNR}, R)$  is given by [14] :

$$E_r(\text{SNR}, R) = \begin{cases} E_{sp}(\text{SNR}, R) & \text{if } R_{rc} < R \leq C \\ E_{rc}(\text{SNR}, R) & \text{if } R_{ex} < R \leq R_{rc} \\ E_{ex}(\text{SNR}, R) & \text{if } 0 < R \leq R_{ex} \end{cases}$$

The boundaries between the regions are as follows. The *Shannon capacity* is  $C \stackrel{\text{def}}{=} \frac{1}{2} \log(1 + \text{SNR})$ . The *critical rate* is  $R_{cr} \stackrel{\text{def}}{=} \frac{1}{2} \log \left( \frac{1}{2} + \frac{\text{SNR}}{4} + \frac{1}{2} \sqrt{1 + \frac{\text{SNR}^2}{4}} \right)$ . The *expurgation rate* is  $R_{ex} \stackrel{\text{def}}{=} \frac{1}{2} \log \left( \frac{1}{2} + \frac{1}{2} \sqrt{1 + \frac{\text{SNR}^2}{4}} \right)$ .

The error exponent in the *sphere packing* region is:

$$E_{sp}(\text{SNR}, R) = \frac{\text{SNR}}{4\beta} \left( \beta + 1 - (\beta - 1) \sqrt{1 + \frac{4\beta}{\text{SNR}(\beta - 1)}} \right) + \frac{1}{2} \ln \left( \beta - \frac{\text{SNR}(\beta - 1)}{2} \sqrt{1 + \frac{4\beta}{\text{SNR}(\beta - 1)}} \right)$$

where  $\beta = 2^{2R}$ . In the *random coding* region:

$$E_{rc}(\text{SNR}, R) = 1 - \beta + \frac{\text{SNR}}{2} + \frac{1}{2} \log \left( \beta - \frac{\text{SNR}}{2} \right) - \frac{1}{2} \log(\beta) - \log(2)R$$

where now  $\beta = 2e^{2R_{cr}}$ . Lastly, in the *expurgation* region:

$$E_{ex}(\text{SNR}, R) = \frac{\text{SNR}}{4} \left[ 1 - \sqrt{1 - 2^{-2R}} \right]$$

It is also possible to show, using some pedestrian algebra, that for all rates  $0 < R < C$ ,  $E_{sp}(\text{SNR}, R)$  coincides with the asymptotic expression of Shannon's sphere packing bound for AWGN [15]. Hence, it is also an upper bound for the reliability function, and thus the bound is tight above the critical rate.

### III. SETUP

Our setup is defined as follows. The feedforward and feedback channels connecting Terminal A to Terminal B and vice versa respectively, are AWGN channels given by

$$Y_n = X_n + Z_n, \quad \tilde{Y}_n = \tilde{X}_n + \tilde{Z}_n.$$

Where  $X_n, Y_n$  (resp.  $\tilde{X}_n, \tilde{Y}_n$ ) are the input and output of the feedforward (resp. feedback) channel at time  $n$  respectively. The feedforward (resp. feedback) channel noise  $Z_n \sim \mathcal{N}(0, \sigma^2)$  (resp.  $\tilde{Z}_n \sim \mathcal{N}(0, \tilde{\sigma}^2)$ ) is independent of the input  $X_n$  (resp.  $\tilde{X}_n$ ), and constitutes an i.i.d. sequence. The feedforward and feedback noise processes are mutually independent.

Terminal A is in possession of a message  $W \sim \text{Uniform}([M])$ , to be described to Terminal B over  $N$  rounds of communication. To that end, the terminals can employ an interactive scheme defined by a pair of functions  $(\varphi, \tilde{\varphi})$  as follows: At time  $n$ , Terminal A sends a function of its message  $W$  and possibly of past feedback channel outputs over the feedforward channel, i.e.,

$$X_n = \varphi(W, \tilde{Y}^{n-1}).$$

Similarly, Terminal B sends function of its past observations to Terminal A over the feedback channel, i.e.,

$$\tilde{X}_n = \tilde{\varphi}(Y^n).$$

**Remark 1.** The dependence of  $\varphi$  and  $\tilde{\varphi}$  on  $n$  is suppressed. In general, we allow these functions to further depend on common randomness shared by the terminals.

We assume that Terminal A (resp. Terminal B) is subject to a power constraint  $P$  (resp.  $\tilde{P}$ ), namely

$$\sum_{n=1}^N \mathbb{E}(X_n^2) \leq N \cdot P, \quad \sum_{n=1}^N \mathbb{E}(\tilde{X}_n^2) \leq N \cdot \tilde{P}.$$

We denote the feedforward (resp. feedback) SNR by  $\text{SNR} \stackrel{\text{def}}{=} \frac{P}{\sigma^2}$  (resp.  $\tilde{\text{SNR}} \stackrel{\text{def}}{=} \frac{\tilde{P}}{\tilde{\sigma}^2}$ ). The ratio between the feedback SNR and the feedforward SNR is denoted by  $\Delta\text{SNR} \stackrel{\text{def}}{=} \frac{\tilde{\text{SNR}}}{\text{SNR}}$ . We implicitly assume that  $\Delta\text{SNR} > 1$ .

An interactive scheme  $(\varphi, \tilde{\varphi})$  is associated with a rate  $R \stackrel{\text{def}}{=} \frac{\log M}{N}$  (in bits) and an error probability  $p_e(N, R)$ , which is the probability that Terminal B errs in decoding the message  $W$  at time  $N$ , under the optimal decision rule. We say that an error exponent  $E(R)$  is achievable if there exists a sequence of interactive coding schemes indexed by  $N$  with rate at least  $R$ , such that  $p_e(N, R) \leq e^{-NE(R)}$ .

### IV. DESCRIPTION OF THE SCHEME

In Subsection II-C we discussed the S-K scheme and described its feedforward transmission as a JSCC of the estimation error. It was assumed that Terminal A knows the estimation error, which is made possible by Terminal B sending back its estimate  $\hat{\Theta}_k$ , and Terminal A in turn subtracting  $\Theta$  to obtain  $\varepsilon_k = \hat{\Theta}_k - \Theta$ . This procedure holds if the feedback is noiseless and fails if the feedback is noisy. In the latter case, it was observed in [6] that the transmission from Terminal B to Terminal A can be regarded as JSCC with side information. Namely, Terminal B wished to convey

$\varepsilon_k$  to Terminal A, but knowing only  $\hat{\Theta}_k = \Theta + \varepsilon_k$  whereas Terminal A knows  $\Theta$  and can use it as side information. The scheme described in [6] used scalar modulo operations and took advantage of the noisy feedback in order to reduce the capacity gap, maintaining the simplicity of the original S-K scheme.

The use of scalar modulo operation benefits from simplicity and low delay, at the price of modulo error which is bounded away from zero. As shown in Subsection II-B, the error probability can be made to approach zero using modulo-lattice operations in the limit of large dimension. This provides motivation to the following modifications of our scalar scheme:

- 1) Replace the scalar interval lattice with a lattice  $\Lambda$  of dimension  $N$ .
- 2) Replace the scalar PAM mapping of the message point  $W \rightarrow \Theta$  with an AWGN block code of the same dimension  $N$ , namely  $W \rightarrow \Theta$ .
- 3) Use block code and lattice error exponents for the analysis of the aggregate error probability incurred by the associated high-dimensional extension of our scalar scheme, where interaction takes place on a block-wise basis.

It should be noted that the feedback operations (i.e. the modulo-lattice operations) requires the knowledge of an entire vector of length  $N$ , and cannot be implemented on the fly. Moreover, the modulo-lattice result requires  $N$  channel uses to be transmitted. To accommodate this inherent delay we use two interlaced block-wise schemes. Having two schemes each using  $K$  rounds requires  $2K$  blocks of length  $N$ . For simplicity, we use double indexing for the blocks. The block index  $l$  is represented by a pair of indices  $(k, j)$  so that  $l = 2(k-1) + j$ . More explicitly, this notation defines

$$\mathbf{X}_k^j \stackrel{\text{def}}{=} [X_{(2(k-1)+j-1)N+1}, \dots, X_{(2(k-1)+j-1)N+N}]$$

We denote the round index by  $k \in [K]$  and the scheme index by  $i \in \{1, 2\}$ . The feedforward of round  $k$  and scheme  $i$  is sent over the block pertaining to indices  $(k, i)$ , and the corresponding feedback is sent over the block pertaining to indices  $(k, i+1)$ .

Let us now give a description of scheme for  $i \in \{1, 2\}$ . The setting of the parameters  $\alpha, \beta_k, \gamma_k$  will be discussed in the sequel. The dither variables  $\mathbf{V}_k^i$  are i.i.d. and uniformly distributed on  $\mathcal{V}_0$ .

(A) Initialization:

**Terminal A:** Map the message  $W^i$  to codeword  $\Theta^i$  using a codebook for AWGN with average power  $P$ .

**Terminal A  $\Rightarrow$  Terminal B:**

- Send  $\mathbf{X}_1^i = \Theta^i$
- Receive  $\mathbf{Y}_1^i = \mathbf{X}_1^i + \mathbf{Z}_1^i$

**Terminal B:** Initialize the  $\Theta^i$  estimate to  $\hat{\Theta}_1^i = \mathbf{Y}_1^i$ .

(B) Iteration:

**Terminal B  $\Rightarrow$  Terminal A:**

- Given the  $\Theta^i$  estimate  $\hat{\Theta}_k^i$ , compute and send in the following block

$$\tilde{\mathbf{X}}_k^{i+1} = \mathbb{M}_\Lambda \left[ \gamma_n \hat{\Theta}_k^i + \mathbf{V}_k^i \right]$$

- Receive  $\tilde{\mathbf{Y}}_k^{i+1} = \tilde{\mathbf{X}}_k^{i+1} + \tilde{\mathbf{Z}}_k^{i+1}$

**Terminal A:** Extract a noisy scaled version of estimation error  $\varepsilon_k^i$ :

$$\tilde{\varepsilon}_k^i = \mathbb{M}_\Lambda \left[ \tilde{\mathbf{Y}}_k^{i+1} - \gamma_n \boldsymbol{\Theta}^i - \mathbf{V}_k^i \right]$$

Note that  $\tilde{\varepsilon}_k^i = \gamma_n \varepsilon_k^i + \tilde{\mathbf{Z}}_k^{i+1}$ , unless a *modulo-aliasing error* occurs.

**Terminal A  $\Rightarrow$  Terminal B:**

- Send a scaled version of  $\tilde{\varepsilon}_k^i$ :  $\mathbf{X}_{k+1}^i = \alpha \tilde{\varepsilon}_k^i$ , where  $\alpha$  is set so that to meet the input power constraint  $P$  (computed later).
- Receive  $\mathbf{Y}_{k+1}^i = \mathbf{X}_{k+1}^i + \mathbf{Z}_{k+1}^i$

**Terminal B:** Update the  $\boldsymbol{\Theta}^i$  estimate  $\hat{\boldsymbol{\Theta}}_{k+1}^i = \hat{\boldsymbol{\Theta}}_k^i - \tilde{\varepsilon}_k^i$ , where

$$\hat{\varepsilon}_k^i = \beta_{k+1} \mathbf{Y}_{k+1}^i \quad (1)$$

is the MMSE estimate of  $\varepsilon_k$ . The optimal selection of  $\beta_k$  is described in the sequel.

- (C) Decoding: After the reception of block  $\mathbf{Y}_K^i$  the receiver decodes the message  $\hat{W}^i(\mathbf{Y}_K^i)$  using an ML decision rule w.r.t. the codebook.

## V. ERROR ANALYSIS AND PARAMETER SETTING

As elaborated above, decoding of the two interlaced schemes produces  $\hat{W}^i(\mathbf{Y}_K^i)$  and an error occurs if either of the decoded messages is not equal to its corresponding sent message  $W^i$ . It is important to note that due to the modulo-lattice operations in the feedback, the additive noise corrupting  $\mathbf{Y}_K^i$  is not Gaussian. However, a Gaussian analysis can be used to bound the error probability as we show herein.

For any  $k \in \{1, \dots, K-1\}$  we define  $E_k^i$  as the event the feedback decoding results in modulo-aliasing error, i.e.

$$E_k^i \stackrel{\text{def}}{=} \left\{ \gamma_n \varepsilon_k^i + \tilde{\mathbf{Z}}_k^{i+1} \notin \mathcal{V}_0 \right\}.$$

We define  $E_K^i$  as the decoding error at the final decoding step

$$E_K^i = \left\{ \hat{W}^i(\mathbf{Y}_K^i) \neq W^i \right\}$$

In order to use the Gaussian analysis we introduce the following upper bound for the error probability:

$$p_e^i \leq \Pr \left( \bigcup_{k=1}^K E_k^i \right). \quad (2)$$

The inequality stems from the fact that a modulo-aliasing error does not necessarily cause a decoding error.

To proceed, we define the *coupled system* as a system that is fed by the same message and experiences the (sample-path) exact same noises, with the only difference being that no modulo operations are implemented at neither of the terminals. Clearly, the coupled system violates the power constraint at Terminal B. However, given the message  $W^i$ , all the random variables in the coupled system are jointly Gaussian, and in particular, the estimation errors  $\varepsilon_k^i$  in that system are Gaussian for  $k = 1, \dots, K$ . Moreover, it is easy to see that the estimation errors are *sample-path identical* between the original system and the coupled system until the first modulo-aliasing error occurs. To be precise we quote [6, Lemma 1]:

**Lemma 1.** Let  $\widetilde{\Pr}$  denote the probability operator in for the coupled process. Then for any  $K > 1$ :

$$\Pr \left( \bigcup_{k=1}^K E_k^i \right) = \widetilde{\Pr} \left( \bigcup_{k=1}^K E_k^i \right).$$

Combining the above with (2) and applying the union bound in the coupled system, we obtain

$$p_e \leq \sum_{i=1}^2 \sum_{k=1}^K \widetilde{\Pr} (E_k^i).$$

Calculating the above probabilities now involves only Gaussian random variable, which significantly simplifies the analysis.

From this step on, we perform an asymptotic exponential analysis. We note that the sums of probabilities are exponentially dominated by the maximal summand, therefore both interlaced schemes  $i \in \{1, 2\}$  are set to be identical, and set the parameters such that all modulo-aliasing error probabilities are the same. Hence

$$p_e \leq 2 \left[ (K-1) \widetilde{\Pr} (E_1^i) + \widetilde{\Pr} (E_K^i) \right] \doteq \widetilde{\Pr} (E_1^1) + \widetilde{\Pr} (E_K^1).$$

Defining  $p_{\text{mod}} \stackrel{\text{def}}{=} \widetilde{\Pr} (E_1^1)$  and  $p_{\text{dec}} \stackrel{\text{def}}{=} \widetilde{\Pr} (E_K^1)$  yields

$$p_e \leq p_{\text{mod}} + p_{\text{dec}}$$

We now set the lattice second moment to equal the feedback power constraint  $\sigma^2(\Lambda) = \tilde{P}$  (and guarantee that this is the feedback transmission power by dithering). The modulo-aliasing error event is the event where

$$\gamma_n \varepsilon_k^i + \tilde{\mathbf{Z}}_k^{i+1} \notin \mathcal{V}_0 \quad (3)$$

By the coupling argument, we can assume for our bounding analysis that the LHS above is Gaussian. The *looseness*  $L$  of the lattice is defined by the power ratio of the RHS and LHS of (3), i.e.,

$$L \stackrel{\text{def}}{=} \frac{\tilde{P}}{\gamma_k^2 \sigma_k^2 + \tilde{\sigma}^2}.$$

By the definitions in Subsection II-A:  $L = \mu(\Lambda) \cdot G(\Lambda)$ . By [13, Theorem 5] there exist lattices that asymptotically attain both  $G(\Lambda) = \frac{1}{2\pi e} + o(1)$  and the Poltyrev exponent, so we can set  $\mu = 2\pi e L + o(1)$ , then  $E_p(\frac{\mu}{2\pi e}) = E_p(L)$  and

$$p_{\text{mod}} \leq e^{-N E_p(L)}.$$

In the next step we send  $\mathbf{X}_{k+1}^i = \alpha \tilde{\varepsilon}_k^i$ , where  $\alpha$  is set so that to meet the input power constraint  $P$ , i.e.  $\alpha = \sqrt{L \frac{P}{\tilde{P}}}$ . The channel output in the next round is thus

$$\mathbf{Y}_{k+1}^i = \alpha \gamma_n \varepsilon_k^i + \alpha \tilde{\mathbf{Z}}_k^{i+1} + \mathbf{Z}_{k+1}^i.$$

Setting  $\beta_{k+1}$  in (1) to the optimal Wiener coefficient, one can easily calculate the evolution of the estimation variance  $\sigma_k^2 \stackrel{\text{def}}{=} \frac{1}{N} \mathbb{E} \|\varepsilon_k^i\|^2$ . We now observe that the channel from  $\boldsymbol{\Theta}^i$  to  $\boldsymbol{\Theta}^i + \varepsilon_k^i$  is in fact a vector of independent parallel AWGN channels each with a noise variance  $\sigma_k^2$ . Namely, after  $K$  rounds, we

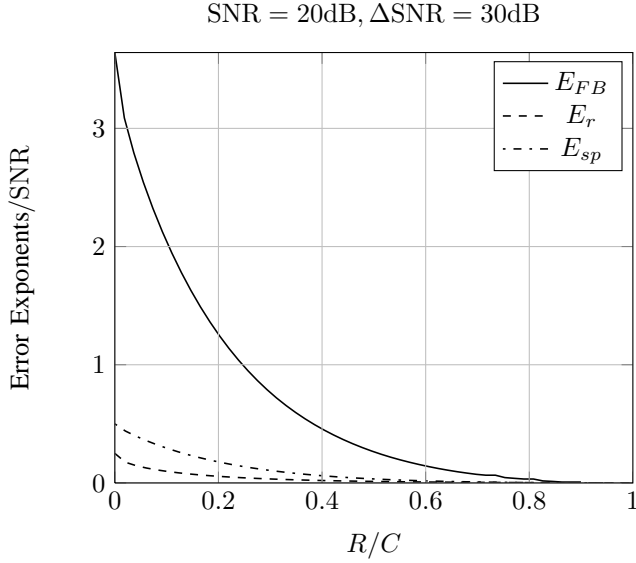


Fig. 1. Error exponents with and without feedback for SNR = 20dB and  $\Delta\text{SNR} = 30\text{dB}$

have  $N$  instances of independent AWGN channels each with SNR given by

$$\text{SNR}_K(L) \stackrel{\text{def}}{=} \text{SNR} \cdot \left( 1 + \text{SNR} \frac{1 - L\tilde{\text{SNR}}^{-1}}{1 + L\Delta\text{SNR}^{-1}} \right)^{K-1}.$$

Therefore, we can now map the message  $W^i$  into  $\Theta^i$  using a Gaussian codebook of block length  $N$  and rate  $K \cdot R$  to obtain

$$p_{\text{dec}} \leq e^{-NE_r(\text{SNR}_K(L), K \cdot R)}.$$

Note that the rate  $K \cdot R$  is chosen such that the overall rate (over  $K$  rounds) is  $R$ . We therefore immediately obtain the following.

**Theorem 1.** *The error probability attained by our suggested interactive scheme is upper bounded by  $p_e \leq e^{-NE_{FB}(R)}$ , where*

$$E_{FB}(R) \stackrel{\text{def}}{=} \max_{K \in \mathbb{N}, L \geq 1} \left\{ \frac{\min\{E_r(\text{SNR}_K(L), KR), E_p(L)\}}{2K} \right\} \quad (4)$$

Note that the division by  $2K$  is due to normalization of the error exponents by the actual code length which is  $2NK$ . The trade-off is now clear: setting the lattice looseness  $L$  to be large reduces  $p_{\text{mod}}$  but also reduces  $\text{SNR}_K(L)$  hence enlarging  $p_{\text{dec}}$ , and vice versa. Due to the monotonicity of  $E_r(\text{SNR}_K(L), K \cdot R)$ ,  $E_p(L)$  in  $L$ , a numerical solution to (4) can be easily found.

## VI. DISCUSSION

Numerical evaluation of  $E_{sp}$ ,  $E_r$  and  $E_{FB}$  for SNR = 20dB and  $\Delta\text{SNR} = 30\text{dB}$  is depicted in Fig. 1. It is clear that in this scenario our scheme improves the error exponents for most rates below capacity.

It is now constructive to give an approximation for high SNR, namely  $\text{SNR} \gg 1$ . It is easy to see that for  $\text{SNR} \gg 1$ :

$$\text{SNR}_K(L) \geq \frac{(\tilde{\text{SNR}} - L)^K}{\Delta\text{SNR} \cdot L^{K-1}} (1 + o(1)).$$

We would now like to set  $E_r(\text{SNR}_K(L), KR) = E_p(L)$  and solve for  $L$ . Assuming both  $E_r$  and  $E_p$  are in their expurgation regions (as shall be verified later) we would like to solve:

$$\frac{1}{4} \frac{(\tilde{\text{SNR}} - L)^K}{\Delta\text{SNR} \cdot L^{K-1}} \eta(KR) = \frac{1}{8} L.$$

where  $\eta(KR) \stackrel{\text{def}}{=} 1 - \sqrt{1 - 2^{-KR}}$ . The solution yields  $L^* = \tilde{\text{SNR}} / (1 + (\frac{1}{2}\eta(KR)\Delta\text{SNR})^{\frac{1}{K}})$ . Plugging it in (4) yields for any  $K > 1$ :

$$E_{FB}(R) \geq \frac{\text{SNR} \cdot \Delta\text{SNR} \cdot \eta(KR)}{16K \left( 1 + (\frac{1}{2}\eta(KR)\Delta\text{SNR})^{\frac{1}{K}} \right)} (1 + o(1)) \quad (5)$$

At  $R = 0$  an optimization on  $K$  is possible, yielding  $K^*$ :

$$K^* = 0.78 \cdot \ln\left(\frac{1}{2}\Delta\text{SNR}\right) \approx 0.18 \cdot \Delta\text{SNR}_{\text{dB}} - 0.54.$$

So either (best of)  $\lceil K^* \rceil$  and  $\lfloor K^* \rfloor$  can be plugged in (5) giving a bound for  $E_{FB}$ . This bound holds as long as this  $K$  and  $L^*$  both satisfy the expurgation region assumptions:  $L > 4$  and  $KR < R_{cr}(\text{SNR}_K(L))$ .

For rates outside this region one can simply use  $\frac{1}{2K} E_r(\text{SNR}_K(L), KR)$  with  $L$  and  $K$  found at the highest rate in the expurgation region.

## VII. ACKNOWLEDGMENT

The authors would like to thank Uri Erez for his help, and Ram Zamir for giving the initial motivation that lead to this work.

## REFERENCES

- [1] C. E. Shannon, "The zero-error capacity of a noisy channel," *IEEE Trans. Inf. Theory*, vol. IT-2, pp. 8–19, Sep 1956.
- [2] J. P. M. Schalkwijk, "A coding scheme for additive noise channels with feedback part II: Band-limited signals," *IEEE Trans. Inf. Theory*, vol. IT-12, pp. 183–189, Apr 1966.
- [3] M. Horstein, "Sequential transmission using noiseless feedback," *IEEE Trans. Info. Theory*, vol. IT-9, pp. 136–143, Jul 1963.
- [4] O. Shayevitz and M. Feder, "Optimal feedback communication via posterior matching," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1186–1222, Mar 2011.
- [5] Y.-H. Kim, A. Lapidoth, and T. Weissman, "On the reliability of Gaussian channels with noisy feedback," in *Proc. 41st Allerton Conf. Communication, Control Computing*, Sep. 2006, pp. 364–371.
- [6] A. Ben-Yishai and O. Shayevitz, "The Gaussian Channel with Noisy Feedback: Near-Capacity Performance via Simple Interaction," in *Proc. 52nd Allerton Conf. Communication, Control Computing*, Oct. 2014, Also available in <http://arxiv.org/abs/1407.8022>.
- [7] M. V. Burnashev and H. Yamamoto, "Noisy feedback improves the gaussian channel reliability function," in *ISIT*, 2014, pp. 2554–2558.
- [8] Y.-H. Kim, A. Lapidoth, and T. Weissman, "Error exponents for the Gaussian channel with active noisy feedback," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1223–1236, March 2011.
- [9] T. Goblick, "Theoretical limitations on the transmission of data from analog sources," *IEEE Trans. Inf. Theory*, vol. IT-11, no. 4, pp. 558–567, Oct 1965.
- [10] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan 1973.
- [11] J. Kochman and R. Zamir, "Joint Wyner-Ziv/Dirty-Paper Coding by Analog Modulo-Lattice Modulation," *IEEE Trans. Inf. Theory*, vol. 55, pp. 4878–4889, 2009.
- [12] R. Zamir, *Lattice Coding for Signals and Networks*, Cambridge University Press, 2014.
- [13] S. Litsyn U. Erez and R. Zamir, "Lattices which are good for (almost) everything," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3401–3416, Oct 2005.
- [14] R. G. Gallager, *Information Theory and Reliable Communication*, New York: John Wiley & Sons, 1968.
- [15] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.*, vol. 38, pp. 611–656, 1959.